International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979, Impact Factor: RJIF 5.72
www.allsubjectjournal.com
Volume 3; Issue 4; April 2016; Page No. 287-291

Heart disease classification and its co-morbid condition detection using WPCA genetic algorithm

[1] Dhivya S, [2] Merlin Mercy E

[1] M.E Student, Department of Computer Science, Sri Krishna College of Technology, Coimbatore Tamil Nadu, India
[2] Asst. Professor, Department of Computer Science, Sri Krishna College of Technology, Coimbatore Tamil Nadu, India

Abstract
In health science, there are several researches and applications are developed using data mining techniques. This paper, contributes an idea to detect the heart disease and its co-morbid conditions along with the risk using data mining techniques. In recent scenario, health issues are huge, due to this nature predicting and classifying into different conditions are very tedious. The field of data mining has involved in those domains to predict and to classify the abnormality along with its risk level. The previous studies have used several features to diagnosis the disease, which has been collected from patients. By applying different data mining algorithms, the patient data is taken for the experiment. The main drawbacks of the previous studies are that need accurate and more number of features. In this paper, a Data mining model has been developed using Weighted Principle Analysis (WPCA) and Genetic algorithm (GA) to improve the prediction accuracy and to investigate the risk level of the disease. The proposed technique helps to the medical domain for predicting heart diseases with its various co-morbid conditions. The study proposed a new classification and prediction scheme for Heart disease data. The system has two main objectives, which are improving diagnosis accuracy and reducing classification delay. The WPCA represents with the effective splitting criteria which has been verified by the genetic algorithm.

Keywords: data mining, heart disease classification, Genetic algorithm, Feature selection

1. Introduction

Use of data mining techniques increases the interpretation and evaluation features in health care domain. Now a day's disease prediction, disease diagnosis and disease condition identification process are very tedious, but this can be ease with the use of data mining techniques. We proposed a new combinatorial data mining algorithm to deal with heart disease diagnosis and risk prediction.

In recent scenario, heart disease is the life threatening problem. Due to many changes in the human life style heart disease is increasing now [1]. Early prediction of those diseases and its risk will helps to achieve maximum recovery. There are number of factors which increase the risk of Heart disease [2] such as hereditary of heart disease, Cholesterol, Diabetes, Smoking, Poor diet, High blood pressure, High blood cholesterol, Obesity, Physical inactivity, Hyper tension, etc., with various test and medical diagnosis, the disease can be identified. In the research area, with huge size of medical conditions and its range the diagnosis process is complicated. To ease the diagnosis process, we proposed a new data mining algorithm. This algorithm predicts and classifies the patient data with risk detection.

There are several co-morbid or types of Heart diseases, which causes different types of issues in the human body. For every condition or disease detection, different types of attributes and rules are needed. The followings are the different types of heart diseases and co-morbid conditions [3].

Coronary heart disease

Coronary heart disease is a type of Heart disease and it is the most common type of heart disease across the world. It is a condition in which plaque deposits block the coronary blood vessels leading to a reduced supply of blood and oxygen to the heart [4].

Arrhythmias

It is associated with a disorder in the rhythmic movement of the heartbeat. The heartbeat can be slow, fast, or irregular. These abnormal heartbeats are caused by a short circuit in the heart's electrical system [5].

Congestive heart failure

It is a condition where the heart cannot pump enough blood to the rest of the body. It is commonly known as heart failure [6].

Congenital heart disease

It also known as congenital heart defect, it refers to the formation of an abnormal heart due to a defect in the structure of the heart or its functioning. It is also a type of congenital disease that children are born with [7].

Cardiomyopathy

It is the weakening of the heart muscle or a change in the structure of the muscle due to inadequate heart pumping. Some of the common causes of cardiomyopathy are hypertension, alcohol consumption, viral infections, and genetic defects [8].

Angina pectoris

It is a medical term for chest pain that occurs due to insufficient supply of blood to the heart. Also known as angina, it is a warning signal for heart attack. The chest pain is at intervals ranging for few seconds or minutes [9].

Myocarditis

It is an inflammation of the heart muscle usually caused by viral, fungal, and bacterial infections affecting the heart. It is an uncommon disease with few symptoms like joins pain, leg swelling or fever that cannot be directly related to the heart [10].

The objectives of this proposal are to diagnosis and predict heart disease and its risk with dynamic feature values in the

healthcare dataset. Improving the quality of the prediction accuracy is also considered. It ensures the quality of values predicted at the time of quality assessment process.

Increasing the classification and prediction accuracy with effective feature reduction technique will also reduce the classification time. In this paper, we propose a new combinatorial data mining algorithm to obtain analytical information of assessing the medical dataset to provide information to the medical team with the likelihood that a patient or user a specific action. The actions comprise of how effective the diseases of the patients are predicted, how the disease are classified using WPCA_GA.

2. Problem Definition

There are several approaches in the literature have been proposed for detection of heart disease. Different supervised machine learning algorithms such as Naïve Bayes, association rule mining algorithms such as Apriori algorithm, Decision algorithm and Neural Network have been used for analyzing the heart disease [1]. The data mining tool such as Weka and Rapid Miner are used for experiment. But the tools and techniques are limited by several factors, such as number of attributes, and failed to find co-morbid conditions and risk level with more number of attributes.

In the existing work, Decision tree classification algorithm has been used to assess the events related to Coronary Artinary Disease. For disease classification, the popular SVM and fuzzy related algorithms are used.

Existing studies do not include some important features. In this paper, several new features are considered in order to increase diagnosis accuracy, while discovering effect of features on CAD. A new feature creation method is used to add three new discriminative features to the patients' records which have a significant impact on prediction ability of the algorithms [11, 12].

3. Proposed System

3.1 Proposed system

We proposed a new combinatorial data mining algorithm for classifying the heart disease and predicting its risk level. The proposed system utilizes the new improved principle component Analysis named as WPCA and modified Genetic Algorithm. Here WPCA reduces the dimensionality problem and finds appropriate feature for finding different co-morbid conditions. And GA used to improve the accuracy by modifying selection process. The followings are the main contributions of the proposed work.

- The system implements a new Genetic based algorithm with the use of effective WPCA dimensionality reduction technique. The system introduces a new Heart disease Classification algorithm with P-GA technique for co-morbid condition detection and its risk level identification.
- This also creates a new advanced feature selection for fast disease classification. The system developed with the intension of high accuracy and less training overhead.
- So the system initially collects and make score for every label, this partially makes an ensemble approach to improve the detection speed.

This analyzed numerous existing classification algorithms and implemented the modified GA and the comparison has been made.

Advantages of the proposed framework

- The proposed method expands the detection of heart disease and its co-morbid conditions and risk level.
- The combinatorial algorithms increase the detection accuracy.
- Improves the accuracy and reduces the false detection and time.

4. Methodologies

This paper presents the analysis of various data mining techniques which can be helpful for disease classification with different types of datasets. This chapter describes the overall process performed in the proposed work. And finally we produced the disease risk level and co-morbid conditions.

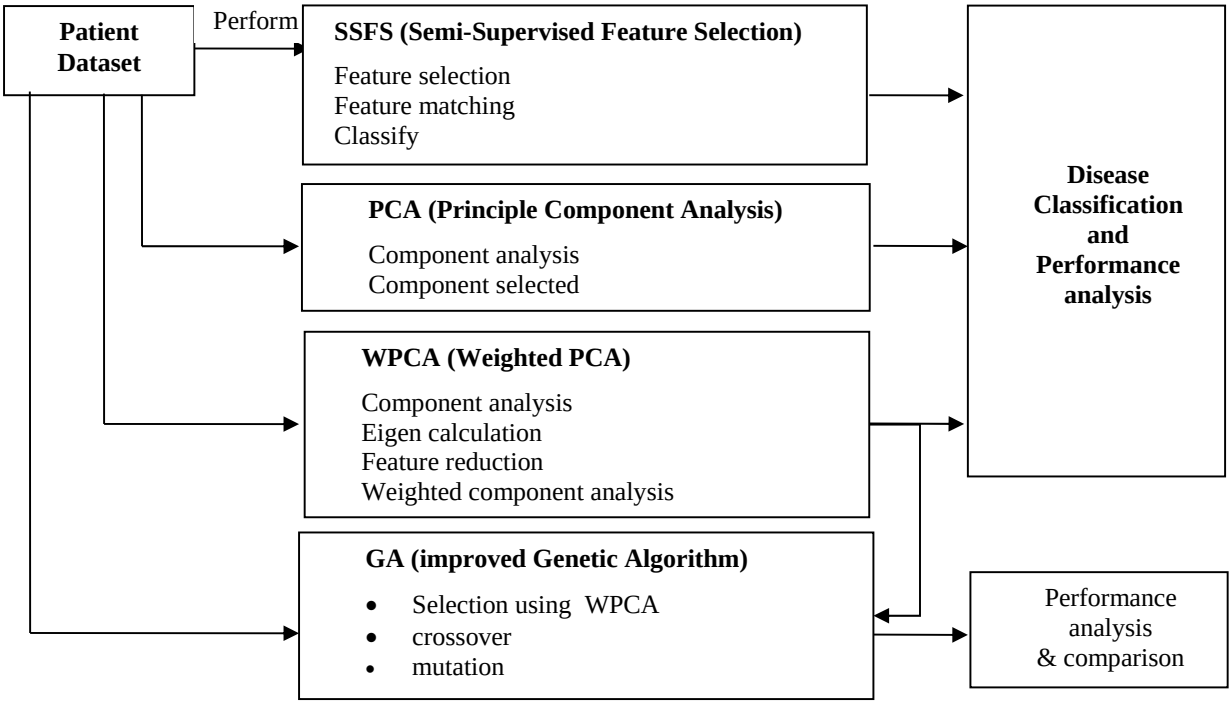

Fig 1: overall process of the proposed work

The above fig 1 shows the overall process of the proposed system, where different algorithms are compared and finally proposed a new improved algorithm for heart disease and its co-morbid conditions detection. Heart disease classification and prediction has been proposed using a renovation algorithm, which is a combination of WPCA (Weighted Principle Component Analysis) and genetic algorithms. Here WPCA is used for feature selection and dimensionality reduction and Genetic algorithm for disease classification and prediction.

The optimized WPCA algorithm has been expanded with the new optimal classification algorithms, which can handle large category dataset more rapidly, accurately and effectively, and keep the good scalability at the same time. The algorithm mainly aims at classified data, but we should disperse the value data in the dealing process. The followings are the advantages of the proposed system.

- WPCA which requires minimum number of training dataset.
- This can be implemented with the following datasets.
  - Healthcare dataset from UCI repository
    Accuracy will be improved using WPCA and genetic

Dataset Description

The data used in this study is the Cleveland Clinic Foundation. Heart disease data set available at http://archive.ics.uci.edu/ml/datasets/Heart+Disease. The data set has 13, 279 attributes. Consequently, to allow comparison with the literature, we restricted testing to these same dataset but increased in attribute size.

Table 1: Dataset description

Dataset	Dataset description
CAD dataset	Coronary Artery disease available at archive.ics.uci.edu/ml/datasets/Heart+Disease With 279 attributes and 500 instances
Statlog Dataset	Available at UCI repository with 13 attributes and 279 instances
SPECTF	Available at UCI repository with 26 attributes and 200 instances

Different dataset is used in the experiment the above table 1 shows the dataset description and its name.

5. Results and Analysis

The first set of experiments is to compare the performance of different existing algorithms for disease classification with different metrics. The experiments are designed so that the different parts of the work could be evaluated. These include the evaluation of the features of the above dataset, the feature selection and also the feature creation methods. To this aim, first the features which were selected by the feature selection method named as WPCA and their importance are discussed. Second, all the four possible combinations of the feature selection and creation methods are theoretically analyzed over the dataset. Finally WPCA and genetic algorithms are implemented this proposed work was implemented using C#.net. The performance of this proposed work WPCA_GA Scheme was compared with the existing algorithms.

The proposed system effectively identifies the disease and its sub types, the sub type which is referred as the percentage of class such as normal and disease. Using combinatorial methods from data mining decision making has been simplified and the proposed work achieved 96.34% accuracy, which is higher than the known approaches in the literature.

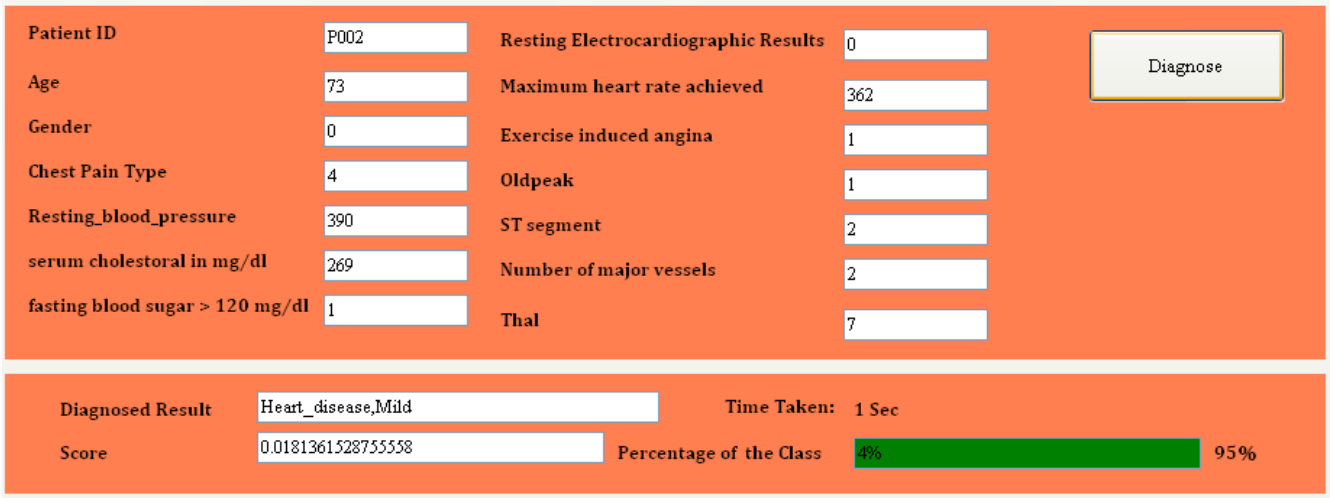

Fig 2: shows the classified result with its risk level and accuracy level.

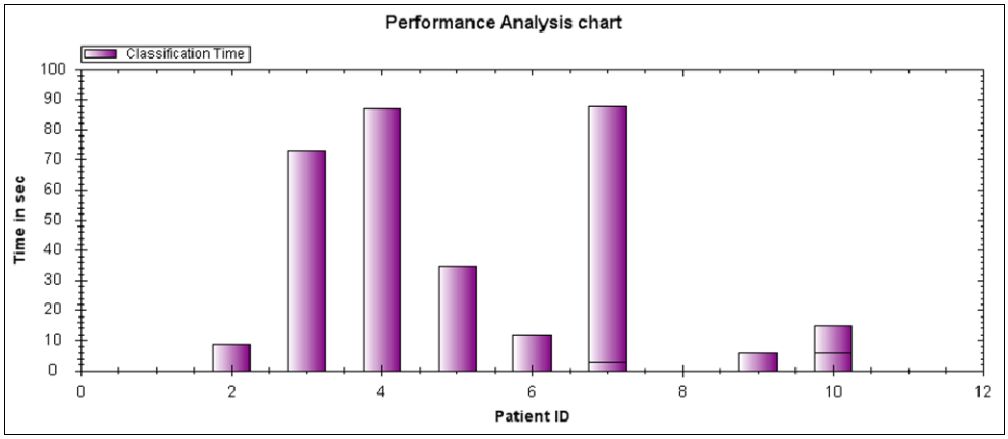

Fig 3: shows the classification time analysis for different patient with different algorithms.

From the above fig 3 the least time taken to diagnosis the disease is our proposed system WPCA_GA. This section evaluates the proposed WPCA_GA with weighted component score based subtype prediction framework in terms of both accuracy and performance. The system applied Hungarian dataset from UCI;

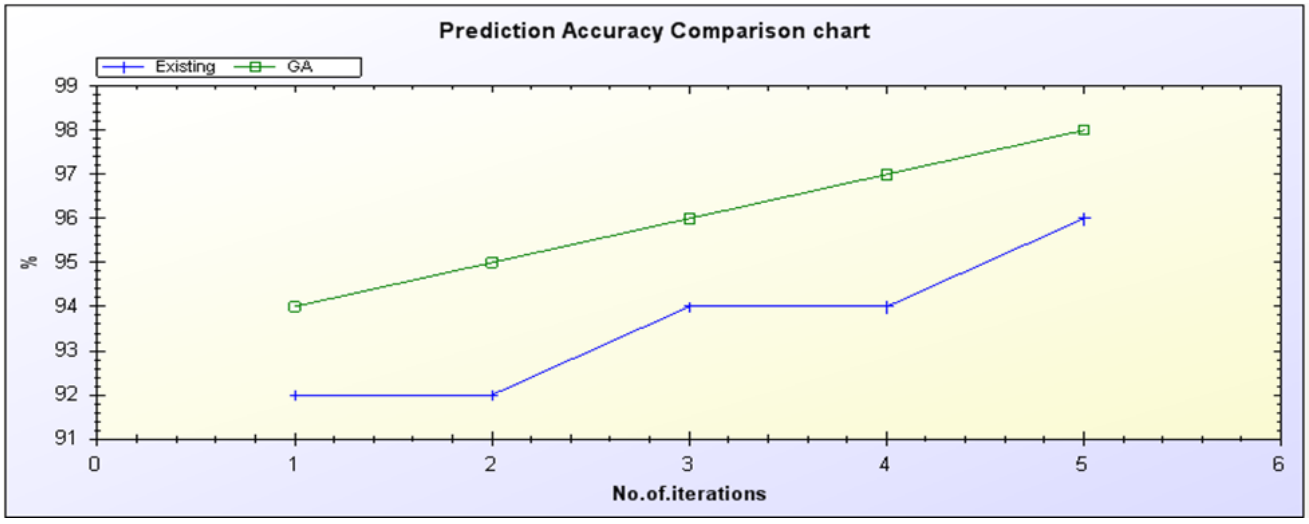

Fig 4: Performance comparison of proposed WPCA_GA with existing PCA approaches based on prediction accuracy.

6. Conclusion and Future Work

The study proposed a new classification and prediction scheme for Heart disease data. The system studied the main two problems in the literature, which are diagnosis accuracy and classification delay. The study overcomes the above two problem by applying the effective enhanced weighted component with genetic algorithm. The WPCA represents with the effective splitting criteria which has been verified by the genetic algorithm. The system effectively identifies the disease and its sub types, the sub type which is referred as the percentage of class such as normal and disease. The experimental results are evaluated using the C#.net. The experimental result shows that integrated extended weighted component with genetic algorithm shows better quality assessment compared to traditional PCA techniques. From the experimental results, the execution time calculated for classification object is almost reduced than the existing system.

7. References

1. Kononenko I. Machine learning for medical diagnosis: History, state of the art and perspective, *Artif. Intell. Med.*, 2001; 23(1):89-109,
2. Magoulas GD, Prentza A. Machine learning in medical applications, *Mach. Learning Appl. (Lecture Notes Comput. Sci.)*, Berlin/Heidelberg, Germany: Springer, 2001; 2049:300-307,
3. Breiman L. Bagging predictors, *Mach. Learning*, 1996; 24(2):123-140,
4. Gordan Kass V. An exploratory Technique for inverstigation large quantities of categorical data Applied Statics, 1980; 29(2):119-127.
5. Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J Stone Classification and Regression Trees. Wadsworth International Group, Belmont, California, 1984.
6. Quinlan JR. Induction of decision trees. Machine Learning, 1986; 1(1):81-106.
7. Zhu Xiaoliang, Wang Jian Yan Hongcan, Wu Shangzhuo. Research and application of the improved algorithm C4.5 on decision tree, 2009.
8. Prof. Nilima Patil, Prof. Rekha Lathi. Comparison of C5.0 & CART Classification algorithms using pruning technique, 2012.

9. Baik S, Bala J. A Decision Tree Algorithm for Distributed Data Mining, 2004.
10. Chauraisa V, Pal S. Data Mining Approach to Detect Heart Diseases, International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2013; 2(4):56-66.
11. Aflori C, Craus M. Grid implementation of the Apriori algorithm Advances in Engineering Software, 2007; 38(5): 295-300.
12. Srinivas K. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques, IEEE Transaction on Computer Science and Education (ICCSE), 2010, 1344-1349,